ABSTRAK

Perkembangan teknologi kecerdasan buatan telah memicu kemunculan *deepfake* yang mengancam keamanan informasi melalui manipulasi wajah yang semakin sulit dibedakan dari citra asli. Penelitian ini bertujuan mengembangkan sistem pendeteksian *deepfake* wajah dengan mengintegrasikan metode *YOLOv11* sebagai detektor wajah dan arsitektur *Xception* sebagai model klasifikasi berbasis *TensorFlow*. Dataset yang digunakan terdiri atas 15.196 citra wajah, yang diperoleh dari beberapa sumber, yaitu *Kaggle, TikTok,* dan *AI Generated*. Seluruh data dikelompokkan ke dalam dua kelas, yaitu *Real* (8.121 citra) dan *Fake* (7.075 citra), kemudian dibagi menjadi data *training* sebanyak 10.450 citra, *validation* sebanyak 2.450 citra, dan *testing* sebanyak 2.296 citra untuk proses pelatihan, validasi, dan evaluasi model. Hasil evaluasi menunjukkan bahwa *YOLOv11* berhasil melokalisasi wajah dengan tingkat keberhasilan 100% pada seluruh dataset yang diuji, termasuk pada kondisi pencahayaan rendah dan tinggi. Model klasifikasi *Xception* mencapai akurasi pengujian sebesar 90,10%. Secara spesifik, nilai *recall* untuk kelas *Fake* mencapai 97,11%, yang menunjukkan sensitivitas tinggi dalam mendeteksi konten manipulatif sehingga meminimalkan kemungkinan *deepfake* diklasifikasikan sebagai citra asli. Nilai *AUC* sebesar 0,98 mengonfirmasi kemampuan diskriminasi model yang berada pada kategori *Excellent*. Analisis visual *Grad-CAM* menunjukkan bahwa model berfokus pada area dahi, pelipis, dan tepi wajah dalam mengidentifikasi artefak manipulasi. Selain itu, aplikasi *Snap Detector* telah divalidasi melalui *black box testing* dengan hasil seluruh fungsi berjalan valid. Namun, sistem menghasilkan kecepatan rata-rata 6,26 *FPS* pada *GPU NVIDIA T4* sehingga masih diperlukan optimasi lebih lanjut untuk memenuhi kebutuhan pemrosesan *real-time* yang lebih optimal.

Kata Kunci: *Deepfake*, *YOLOv11*, *Xception*, klasifikasi, *Computer vision*.

ABSTRACT

The rapid advancement of artificial intelligence technology has led to the emergence of deepfake content, which poses a significant threat to information security through facial manipulations that are increasingly difficult to distinguish from authentic images. This study aims to develop a deepfake face detection system by integrating the YOLOv11 method as a face detector and the Xception architecture as a TensorFlow-based classification model. The dataset consisted of 15,196 facial images collected from multiple sources, including Kaggle, TikTok, and AI-generated content. The dataset was categorized into two classes, namely Real (8,121 images) and Fake (7,075 images), and then divided into 10,450 training images, 2,450 validation images, and 2,296 testing images for the training, validation, and evaluation processes. In the proposed system, YOLOv11 was employed to detect and localize facial regions, followed by a preprocessing stage before the cropped face images were classified by the Xception model as either Real or Fake. The evaluation results demonstrated that YOLOv11 successfully localized facial regions with a 100% detection rate across all tested datasets, including images captured under both low- and high-lighting conditions. The Xception classification model achieved a testing accuracy of 90.10%. Specifically, the recall for the Fake class reached 97.11%, indicating high sensitivity in detecting manipulated content and minimizing the likelihood of deepfake images being misclassified as authentic. Furthermore, the model achieved an AUC score of 0.98, indicating an Excellent level of discrimination capability. Grad-CAM visualization confirmed that the model focused primarily on the forehead, temples, and facial boundaries to identify manipulation artifacts. In addition, the Snap Detector application was successfully validated through black-box testing, with all system functions operating correctly. However, the system achieved an average processing speed of 6.26 FPS on an NVIDIA T4 GPU, indicating that further optimization is required to meet the requirements of real-time processing. Overall, the integration of YOLOv11 and Xception demonstrated promising performance for deepfake face detection and has the potential to support digital facial authenticity verification.

Keywords: *Deepfake, YOLOv11, Xception, Classification, Computer vision.*